DOCUMENT RESUME

ED 320 931 TM 015 150

AUTHOR Huberty, Carl J.; Wisenbaker, Joseph M.
TITLE Variable Importance in MANOVA and DA.
PUB DATE Apr 90
NOTE 43p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Comparative Analysis; *Discriminant Analysis; *Multivariate Analysis; Predictor Variables; *Research Methodology
IDENTIFIERS Bootstrap Hypothesis; Rank Order; *Variables (Mathematics)

ABSTRACT

Some interpretations of relative variable importance in the contexts of multivariate analysis of variance (MANOVA) and discriminant analysis (DA) are presented. Some indices potentially useful for the interpretations are presented, and the assessment of variable importance is illustrated using real data sets. Both descriptive discriminant analysis (DDA) and predictive discriminant analysis (PDA) are considered. Methods for numerically rank-ordering variables in DDA and PDA are described. Two real data sets are used to illustrate the applications for the ranking methods with the preferred indices of relative variable importance. One set consists of three groups of sizes 35, 81, and 37; the other of three groups of sizes 36, 27, and 23. A resampling approach, the bootstrap, was used to partially assess ranking method effectiveness. Taken as a whole, the examples warn against naively interpreting strict rankings of statistics related to variable importance. Ten tables present data from the applications of the various methods. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CARL J. HUBERTY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED320931

Variable Importance in MANOVA and DA

Carl J Huberty Joseph M. Wisenbaker

University of Georgia

325 Aderhold Hall
University of Georgia
Athens, GA 30602

Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1990.

TM015150

BEST COPY AVAILABLE

Variance Importance in MANOVA and DA

Introduction	1	
Descriptive Discriminant Analysis	2	
Contribution to Construct Definition		3
Contribution to LDF Scores		5
Contribution to Grouping Variable Effects		7
Conclusion		10
Two-Way Layout		10
Contrasts		11
Variable Selection		12
Ranking of Variables		12
DDA Examples		15
Predictive Discriminant Analysis	18	
Ranking of Variables		20
PDA Examples		21
Comments	23	
References	26	

Variable Importance in MANOVA and DA

Although statisticians do not relate much at all to the problem of relative variable importance in data analysis (exceptions are Kruskal, 1984; Kruskal & Majors, 1989; Mosteller & Wallace, 1963; Schaafsma & Van Vark, 1979), implications of the use of the concept are very often made by applied social and behavioral science researchers in the analysis of data on multiple response variables. It is fairly common to read statements such as "the most relevant variable was...," or "...was the strongest predictor," or "the criterion variable was primarily a function of...," or "...was clearly the most valid predictor," or "the most promising variables were...," or "...had the most direct influence," or "...had the greatest impact of all of the variables," or "...is a primary determinant," or "...was the most important variable." References to relative variable importance or variable contribution in the applied literature are ubiquitous.

Authors of empirical journal articles do not always make clear what is precisely meant by words such as "important," or "powerful," or "strong," or "promising" when discussing relative variable contribution. That is, variable contribution to what? The purposes of this paper are to present some interpretations of relative variable importance in the contexts of multivariate analysis of variance (MANOVA) and discriminant analysis (DA), to

present some indices potentially useful for the interpretations, and to illustrate the assessment of variable importance using real data sets.

In the next section the variable importance issue is discussed in the context of MANOVA and descriptive discriminant analysis (DDA). This is followed by a discussion of the variable importance issue in predictive discriminant analysis (PDA). Methods for numerically rank-ordering variables in DDA and PDA are described. Two real data sets are utilized to illustrate the applications of the ranking methods with our preferred indices of relative variable importance. A resampling approach, the bootstrap, is employed to partially assess ranking method effectiveness.

Descriptive Discriminant Analysis

Only a few methodological articles were found that have addressed the issue of relative variable importance in the context of MANOVA/DDA. Two articles (Eisenbeis, Gilbert, & Avery, 1973; Karson & Martell, 1980) are restricted to the two-group situation, and one article (Van Vark, 1976) deals with data analysis in biometry.

In MANOVA and DDA, there are at least three ways of viewing relative variable importance ("variable" here refers to outcome variable):

1) Relative contribution to latent construct definition;
2) Relative contribution to linear discriminant function (LDF) scores; and

3) Relative contribution to grouping variable effects.

These three are included because of their potential usefulness or because they have been utilized by applied researchers. The first and second views of variable importance are quite popular with applied researchers, particularly the second view.

The initial discussion of these views of relative variable importance will pertain to a one-way layout design; i.e., a one-factor design with k levels, and p outcome variables. Initially, also, focus will be on the one-way MANOVA omnibus hypothesis tested when the condition of equal population covariance matrices is judged tenable. Later, mention will be made of variable importance in the contexts of two-way MANOVA and multivariate contrast analyses.

Contribution to Construct Definition

The approach to identifying/defining/naming constructs that underlie grouping variable effects taken here is that of examining correlations between each outcome variable and each linear discriminant function (LDF). [In the one-way layout context, the maximum number of LDFs to consider is m, the minimum of p and k-1.] These correlations are often called structure r's or loadings; only within-groups correlations are considered here. The variables that correlate highly with an LDF are conceptually "put together" to arrive at a name for the construct -- this approach is very common in identifying constructs in common factor analysis.

In the context of descriptive discriminant analysis at least, the use of structure r's for the purpose of ordering variables is problematic. The numerical signs of the r values may be ignored; squaring or using absolute values would be appropriate. Herein we consider r^2 values. With two groups of experimental units -- k=2 -- there is a single LDF and, therefore, one set of r^2 values. It turns out that the p r^2 values are proportional to the p univariate F (or Student t) values -- see Huberty (1972) for proof. Thus, a variable ordering based on r^2 values would be identical to an ordering based on univariate F or Student t values. An ordering based on r^2 values would in essence, then, ignore the intercorrelations of the outcome variables. This is particularly disturbing to those of us who would prefer to stay in a <u>multivariate</u> mode while making interpretations of a <u>multivariate</u> analysis (Huberty & Morris, 1989).

Consider now an analysis involving k > 2 groups. On the surface, this appears to be a more complicated situation. There are m sets of structure r values; one set of p r values for each of the m LDFs. For a given variable, how do we "combine" the m r^2 values? One possibility is to simply sum the m r^2 values to arrive at a variable ordering index -- this is like a <u>communality</u> in factor analysis (see Pruzek, 1990). This linear composite implies equal weighting of the r^2 values across the m LDFs; i.e., the m LDFs are thusly considered equally "important." A more reasonable ordering index is a linear composite using the eigenvalue (λ_j) of $E^{-1}H$ associated with the jth LDF as a weight for the corresponding

r^2 value -- E and H are the error and hypothesis sums of squares and cross-products matrices, respectively. That is, this latter ordering index is

$$c_i = \sum_{j=1}^{m} \lambda_j r_{ij}^2$$

A rationale for associating a "weight" of λ_j with the jth LDF is that λ_j reflects the ratio of between-group to within-group variability with respect to scores on the jth LDF (Tatsuoka, 1988, p. 213). That is, a λ_j value reflects a proportion of variation in the p-variable system accounted for by the jth LDF. An argument might be made, then, that if a variable yields a high c_i value, then it is contributing relatively more to the definitions of the m latent constructs than are variables with lower c_i values. It turns out, similar to the k=2 case, that the p c_i values are proportional to the p univariate F values -- see Rencher (1988) for proof. Thus, the use of the above composite as a variable ordering index in DDA -- a multivariate context -- is a dubious practice. It is more reasonable to us to look elsewhere for an index.

Contribution to LDF Scores

Linear discriminant functions (LDFs) are linear composites of outcome variables, the variable weights being determined so as to maximize group separation (in a one-way layout) with respect to composite scores. That is, for a given data set, the ANOVA F value based on LDF scores is greater than an F value based on scores on any other linear composite. To determine relative variable importance when there is only one LDF, then, one might examine the

relative magnitudes of the weights used to form the LDF, provided measures on the p variables are commensurable. If all p variables are not measured using the same scale, then variable weights may be determined that are applicable to variable standard scores, often in z-score form. It should be pointed out that the SAS package differs from BMDP and SPSS-X in how the "standardized weights" are calculated -- see Mueller & Cozad (1988). The use of within-groups variances (a la BMDP and SPSS-X) is generally preferred to the use of total-group variances (a la SAS) in transforming raw weights to standardized weights.

Supposedly the rationale for using the standardized weight as a variable ordering index is that a large weight, in an absolute value sense, indicates that the associated variable contributes greatly to the LDF score for each experimental unit. And it is with respect to these LDF scores that the groups are maximally separated.

There is little doubt that the most popular approach to ordering outcome variables in the context of MANOVA and DDA is to use these so-called standardized LDF weights. In a two-group situation where there is only one LDF, the variable with the largest absolute standardized weight is considered the most important variable. If one were to use the standardized LDF weight, b_i^*, as an ordering index, it would be desirable to take into consideration the sampling fluctuation of b_i^*. This may be accomplished by dividing $(b_i^*)^2$ by r^{ii}, the ith diagonal element of

the inverse of the pxp within-groups outcome variable correlation matrix (see Pedhazur, 1982, pp. 81, 207). This index,

$$u_i = (b_i^*)^2/r^{ii},$$

is an estimate of the decrease in group separation when Y_i is deleted (see Urbakh, 1971, p. 532). As such, then, ordering variables according to u_i values when k=2 is equivalent to ordering variables according to F-to-remove values, $F_{(i)}$, discussed in the next section. [See Huberty and Smith (1982, pp. 420-421) for greater detail on u_i.]

If the number of LDFs, m, is greater than one, variable ordering via standardized weights is problematic. That is, how are the m standardized weights for a given variable combined to arrive at an ordering index value? One approach is to follow a strategy similar to that discussed in the subsection on Contribution to Construct Definition to get c_i values. Eisenbeis and Avery (1972, p. 70) suggest that one might multiply a variable's standardized weight by the eigenvalue (of $E^{-1}H$) associated with each LDF to arrive at an ordering index value. Since the signs of the b_i^*'s are to some extent arbitrary, one needs to consider either the absolute value of b_i^*, or $(b_i^*)^2$. No matter which of these is used, the sampling fluctuation of b_i^* is ignored. Thus, we question basing a variable ordering index on b_i^* values.

Contribution to Grouping Variable Effects

One way to interpret the third view of relative variable contribution is that of contribution to the variation shared by

grouping variable effects and the collection of outcome variables. The shared variation may be determined using a multivariate eta-squared value (the complement of a Wilks lambda) for a particular set of effects. [Note: The generic expression, "grouping variable effects," is used to cover MANOVA main effects, interaction effects, and contrast effects.] The relative contribution of Y_i to grouping variable effects in a one-way context may be assessed by answering the following questions:

a. How well can we do without Y_i?,

b. What does Y_i contribute in addition to other variables?, or

c. What does Y_i contribute, controlling for the other variables?

[The idea in the last question was expressed in a multiple regression context about three decades ago by Blalock (1961) -- unfortunately, the idea has not been taken seriously by applied researchers over the years.] The first question may be answered directly by determining the extent of group separation when Y_i is deleted from the original set of p outcome variables. If there is very little separation when Y_i is deleted (i.e., a large Wilks lambda value results), then Y_i is judged to be relatively important. One way, then, of ordering the variables is to conduct p analyses, each utilizing p-1 outcome variables. The SPSS-X computer package, however, yields an index value for each variable in a single analysis. In the last step of a DISCRIMINANT printout is a WILKS' LAMBDA for each variable [$\Lambda_{(i)}$]. For variable Y_i, the

$\Lambda_{(i)}$ value is the Wilks lambda value using all variables except Y_i. Therefore, in the sense of the first question, variables may be ordered with respect to their contribution to group separation by simply ordering the p $\Lambda_{(i)}$ values.

An ordering of the variables in the sense of the other two questions may also be otained from package printouts. The SPSS-X DISCRIMINANT and BMDP 7M programs output an F TO REMOVE value for each variable. Such an $F_{(i)}$ value is a transformation of a partial lambda that reflects the separation yielded by Y_i in addition to the other p-1 variables $[\Lambda(Y_i|\underline{Y}_{(i)})]$. Equivalently, $F_{(i)}$ is the test statistic used in an analysis of covariance with Y_i as the outcome variable and the other p-1 variables as covariates -- this is an example of the "additional information" idea of the Indian statistician, C. R. Rao. The same $F_{(i)}$ values are outputted by the SAS STEPDISC program where they are labeled simply as "F." An $F_{(i)}$ value, then, is a transformation of a reasonable index of relative variable importance (partial lambda) and involves a consideration of sampling variability and, in turn, represents a reasonable variable ordering index. A high $F_{(i)}$ value for Y_i indicates that the separation among the groups would be drastically decreased if Y_i were deleted. Thus, a variable with a higher $F_{(i)}$ value, relative to those for the other variables, would be judged to be more important. The STEPDISC program also yields a PARTIAL R**2 $[R^2_{(i)}]$ value for each variable -- this is simply the complement of the partial lambda for Y_i. Large $R^2_{(i)}$ values indicate relatively important variables.

From the definitions of $\Lambda_{(i)}$, $F_{(i)}$, and $R^2_{(i)}$, it is perhaps clear that variable orderings yielded by their numerical values will be identical. Therefore, in the subsequent examples discussed, only the $F_{(i)}$ values which are common across all the packages will be presented.

Conclusion

Our preference of an interpretation of, and index for, relative variable importance in the context of a one-way MANOVA and associated DDA is perhaps clear from the above discussion. We prefer to view variable importance in terms of relative contribution to grouping variable effects. And our preferred ordering index is the "F-to-remove," $F_{(i)}$.

Two-Way Layout

The initial omnibus hypothesis considered for data based on a two-factor (A-by-B) design is the interaction hypothesis. If the interaction effects are generalizable, then one might be interested in determining the relative contribution of the outcome variable to the resultant interaction effects; i.e., the relative contribution of Y_i to the variation shared by the AB interaction and the outcome variable set. Obtaining the $F_{(i)}$ index values via the popular program packages is somewhat problematic. Only through the SPSS-X DISCRIMINANT program can the $F_{(i)}$ values be otained directly. Also in the presence of interaction effects one may test A effects for each level of B (or, vice versa). The $F_{(i)}$ values associated

each set of A effects may also be obtained, with some effort, using SPSS-X DISCRIMINANT. Of course, for either the interaction effects or the A effects at each level of B, multiple analyses may be run by dropping one outcome variable for each run. If the interaction effects are not generalizable, then $F_{(i)}$ values are obtainable via SPSS-X DISCRIMINANT to order the outcome variables in terms of contribution to the main A effects and to the main B effects.

From a substantive theoretical viewpoint, multiple variable orderings may very well be informative. That is, it may be informative to learn that some variables contribute heavily to, say, A effects, while other variables contribute heavily to B effects. Or, it may be informative to learn that some variables contribute heavily to A effects at level B_2, while other variables contribute heavily to A effects at level B_3. [The latter conclusion may lend to a partial "explanation" of interaction effects.]

Contrasts

A general strategy for studying multivariate group contrasts is presented by Huberty and Smith (1982). The details of how to accomplish the ordering of outcome variables according to $F_{(i)}$ values for contrast effects are discussed by Huberty, Chou, and Benitez (1990). Here too, it may be substantively informative to learn that some variables contribute heavily to one set of contrast effects, while other variables contribute heavily to another set of contrast effects.

Variable Selection

There may be times when a researcher is interested in determining which subsets of variables are important and which are not. This pertains to the variable selection problem. The many methods proposed for the selection of "best" subsets of outcome variables in the MANOVA/DDA context are discussed by McKay and Campbell (1982). The all-possible-subsets methods of McCabe (1975) has been advocated elsewhere (Huberty, 1984, 1989), with cautions, as a preferred method of selecting best variable subsets. According to this method, the best subset of a given size (q) is that subset which yields the smallest Wilks lambda value. This is equivalent to concluding that the best subset of size q is that subset for which the complementary subset of size p-q yields the largest Wilks lambda value of all subsets of size p-q. Thus it may be concluded that the variable-ordering approach that we prefer would be "consistent" with preferences for the all-possible-subsets method of variable selection.

Ranking of Variables

We do not recommend that a variable ordering be determined by simply assigning ranks to the $F_{(i)}$ values for the p variables. That is, something other than a naive 1, 2, ..., p ranking (call this Method N) should be considered. Suppose $F_{(8)} = 5.39$ is the largest obtained $F_{(i)}$ value, while $F_{(3)} = 5.28$ is the second largest. To assign Y_8 a rank of 1 and Y_3 a rank of 2 seems to be placing an

inordinate degree of confidence in the stability of those values. Therefore, we suggest two alternative approaches to ordering the variables.

One alternative ranking procedure may be described as follows. First, order the p F-to-Remove values from highest to lowest. Then, find the p-1 consecutive F differences, d_i. The next step is to determine if each d_i is large enough to suggest a "cut-point" in the set of F values. To arrive at a rough indication of the variability of the differences, the variance of the central F distribution (Hogg & Tanis, 1977, p. 214),

$$\sigma_{F(v,w)} = 2w^2(v + w - 2)/v(w-2)^2(w-4),$$

is used to arrive at a divisor for each d_i, $S_d = .5\sqrt{2\sigma_F}$. [It is recognized that: 1) a central F referent distribution is not strictly legitimate; and 2) the F-to-Remove statistics are not independent.] The criterion for determining a "large" d_i and, therefore, a cut-point in the set of Fs is $d_i \geq S_d$. The rationale for using the factor of .5 is to err on the side of finding too many cut-points. Ranks are assigned, then, in the usual way of a "1" for the highest F-to-remove, etc., except that mid-ranks are assigned if there are multiple Fs above the highest cut-point, between cut-points, or below the lowest cut-point. Call this Method D.

The second alternative ranking procedure may be described as follows. First calculate the value of σ_F using the above expression. Then, identify the variable with the "best" index value, the variable with the largest $F_{(i)}$ value. Say $F_{(1)}$ is largest.

Determine $F_{(1)} - \sigma_F$ Variables with index values greater than $F_{(1)} - \sigma_F$ are assigned the same rank (using the mid-rank). Suppose $F_{(2)}$ is the largest index value that is less than $F_{(1)} - \sigma_F$. Determine $F_{(2)} - \sigma_F$; variables with index values greater than $F_{(2)} - \sigma_F$, and less than or equal to $F_{(2)}$, are assigned the same rank. Continue until all p variables are ranked. Call this Method S.

A problem to be addressed pertains to the extent Method D and Method S would yield reasonable variable orderings. To address this problem in the examples used in this paper, the resampling approach of bootstrapping was employed. For each dataset of size N, 100 bootstrap samples of size N were selected without consideration of group membership. For each bootstrap sample, variables were ranked according to the numerical value of F-to-remove; i.e., Method N was used with each bootstrap sample. Two results of the resampling for a given data set were considered. One is the count, for each variable, of the number of bootstrap samples that yielded a given rank (according to Method N). The second result of interest involves a statistical assessment of all pairwise differences of bootstrapped variable mean F-to-remove values (using "dependent t tests"). For example, is the mean of the 100 bootstrap Variable 1 F-to-remove values statistically different from the mean of the 100 Variable 2 F-to-remove values? The two results (variable-rank counts, and pairwise variable F-to-remove differences) are used in an attempt to assess the appropriateness of variable rank-ordering by Method D and Method S.

DDA Examples

Two real data sets are considered. Both sets are based on a one-factor independent-groups design. No substantive interpretations of analysis results will be attempted. Each data set was reviewed for potential outlier observation vectors. Outliers may be identified by examining the "typicality probabilities" associated with each unit (Huberty & Wisenbaker, in press) -- a typicality probability is the estimated probability of a unit having its observation vector, given the unit is assigned to the designated group. If the group membership (posterior) probability for an individual unit is highest for, say, Group 1, but that unit's vector is far (in a probabilistic sense) from the centroid of Group 1, then the unit is a potential outlier. The condition of equal population covariance matrices was also examined (through an approximate F transformation of the Box M statistic) with the conclusion that the condition has been met to a reasonable extent for both data sets. The within-groups outcome variable correlation matrix was examined for possible variable redundancy -- no redundancies were, in our judgment, present in either of the data sets.

All data analyses done for this paper were accomplished via computer programs written by the second author using SAS IML. Algorithm accuracy was assessed against the output of SAS PROC STEPDISC.

Data Set ONE

This data set has three groups of sizes 35, 81, and 37 (N = 153), with measures on 13 response variables. Descriptive information is given in Table 1. The F-to-remove values, along with the variable rank-orderings according to the three methods, are presented in Table 2. The variable-rank counts arising from the bootstrap samples are given in Table 3. It is clear that there

---

Insert Tables 1, 2, 3 About Here

---

is considerable variability in the naive rank assignments for all variables except for V11 and V12. These two variables, as assessed by all three ranking methods, are clearly the most important variables (according to our interpretation of relative importance). One might conclude, from Table 3, that V1 and V8 are two of the least important variables; this conclusion is consistent with the ranking yielded by Method S. Some 78 t tests were conducted to test for differences of pairs of F-to-remove means based on the 100 bootstrap samples. Results of the tests are summarized in Table 2 with the vertical lines. Thus, the mean Fs for V12, V11, V4, and V5 are all different (P at most equal to .001) from each other and from means of all other variables. Variables V9, V3, ..., V10 are, according to the t tests, all "comparable," while V8 and V1 are significantly less important (on the average) than all other variables.

There is a fair bit of agreement among the t test results and the rankings yielded by Method D and Method S. In this case there isn't a clear basis for choosing between the two non-naive ranking methods -- clearly Method N is inappropriate since it would be questionable to conclude that, for example, V9 is more important than V2.

Data Set TWO

This data set has three groups of sizes 36, 27, and 23 (N=86), with measures on five response variables. Descriptive information is given in Table 4. The F-to-remove values, along with the variable rank-orderings, are presented in Table 5. The variable-rank counts based on the bootstrap samples are given in Table 6. This latter table reflects the instability of the naive variable

---

Insert Tables 4, 5, 6 About Here

---

ranks across repeated sampling. The results of the 10 pairwise Student t tests are reflected simply with the vertical line next to the F-to-remove values in Table 5. That is, the mean bootstrap Fs for variables V4, V1, and V2 are not statistically different -- all three P values were .04 or greater. The mean bootstrap F for V5 was, however, found to be statistically larger than the means for the other four variables, and V3 smaller than the others.

Our interpretation of these results (presented in Tables 5 and 6) is that the naive method of ranking variables should not be trusted. Furthermore, both Method S and Method D may be viable

ranking options. So, how might the variables be ordered? Our judgment would be to go with Method D.

Predictive Discriminant Analysis

Whereas descriptive discriminant analysis (DDA) pertains to the description of effects of the grouping variable(s), predictive discriminant analysis (PDA) pertains to the prediction of group membership for experimental units. That is, PDA pertains to classification of units into previously established groups. Whereas, too, an index of the existence of grouping variable effects in DDA is the Wilks lambda, an index of prediction accuracy in PDA is a proportion of correct classification. [An extended discussion of DDA-PDA distinctions is given by Huberty (1984).]

The details of potential classification rules are discussed elsewhere (James, 1985). Generally speaking, there are two classes of classification rules, linear and quadratic, both of which involve considerations of estimates of prior probabilities of group membership and of misclassification costs. Whatever the choice of rule, most classification results may be obtained via programs from the three packages, BMDP, SAS, and SPSS-X -- details of obtainable information are discussed by Huberty and Wisenbaker (in press).

There is a number of ways of assessing the predictive accuracy of a set of predictors -- see Huberty, Wisenbaker, and Smith (1987). The index of predictive accuracy of interest here is the percent of correct classifications, or, a "hit rate." No matter

which type of classification rule is used, linear or quadratic with or without additional considerations, one may conduct either an internal or external classification analysis. An internal analysis is one in which the rule built on a set of objects/subjects is used to classify the very same set of objects. An external analysis is one in which the rule is built on one set of objects and is applied to a different set of objects. Generally, we favor the use of an external classification analysis -- in particular, the leave-one-out method. Of course, a researcher can focus on predictive accuracy across all criterion groups, or on the accuracy for any one particular group. For purposes of the current discussion, we will focus on total-group hit rate for the complete set of predictors, or on predictor subsets.

Relative variable importance in the context of PDA, then, will be determined according to predictive accuracy. The interpretation of "importance" taken here is the very same as that in the context of MANOVA/DDA: a predictor is relatively important if, when deleted, the total-group hit rate for the remaining predictors is considerably less than the hit rate yielded by the complete set of predictors. Again, we want to answer the question, "How well can we do without Y_i?" That is, "What does Y_i contribute to total-group classification accuracy in addition to the other predictors?" The answer to this question may be obtained by multiple runs (as many as there are predictors) of a computer program. Although an external analysis is favored for the purpose of estimating predictive accuracy, we conjecture that for the purpose of

predictor ordering, an internal analysis will suffice. [Therefore, either of the SAS DISCRIM or the SPSS-X DISCRIMINANT programs may be employed.] We let $H_{(i)}$ denote the total-group internal hit rate yielded by all predictors except the ith one; the smaller $H_{(i)}$, the more important is predictor Y_i.

Ranking of Variables

Numerical orderings of predictor variables in the PDA context may be determined by applying procedures similar to those applied in the MANOVA/DDA context. One alternative (to Method N) starts by finding the expected number of total-group hits:

$$e = \Sigma e_g = \Sigma q_g n_g,$$

where q_g is the estimated prior probability membership for Group g. The hit rate, $H_{(i)}$, associated with predictor Y_i is transformed as follows:

$$Z_{(i)} = (NH_{(i)} - e)/\sqrt{e(N-e)/N},$$

where $N = \Sigma n_g$. These $Z_{(i)}$ values are ordered from lowest to highest, and differences, d_i, of consecutive $Z_{(i)}$ values are determined. The criterion for determining a cut-point in the $Z_{(i)}$ values is $d_i \geq .5\sqrt{2}$. Call this Method DZ.

The second alternative ranking procedure starts with transforming the $H_{(i)}$ values to $Z_{(i)}$ values as above, and ordering the $Z_{(i)}$ values from lowest to highest. The "best" predictor is the one which is associated with the lowest $Z_{(i)}$ value. Say $Z_{(1)}$ is lowest. Determine $Z_{(1)} + .5$; predictors with $Z_{(i)}$ values less than $Z_{(1)} + .5$ are assigned the same rank. Suppose $Z_{(2)}$ is the smallest value greater

than $Z_{(1)} + .5$. Determine $Z_{(2)} + .5$ and greater than or equal to $Z_{(2)}$ are assigned the same rank. Continue until all p predictors are ranked. Call this Method SZ.

Again, we want to study the reasonableness of Methods DZ and SZ. As with DDA, we utilized a bootstrap approach. For each bootstrap sample values of $H_{(i)}$ -- the total-group hit rate using all predictors except Y_i -- were calculated. Using these $H_{(i)}$ values, the predictors were ranked via Method N -- the naive method. Again, just as for the DDA situation, predictor-rank counts and tests of pairwise predictor $Z_{(i)}$ differences were used in the assessment.

PDA Examples

The same data sets used for the DDA examples are utilized in the PDA context.

Data Set ONE

Descriptive information for this data set is given in Table 1. The total-group hit rate using all 13 predictors is 133/153 = .869. The hit rates, $H_{(i)}$, obtained by deleting each predictor, one at a time, are given in Table 7. Also reported in Table 7 are the transformations of each $H_{(i)}$ to $Z_{(i)}$, along with rankings of the predictors according to the three methods. The application of the naive method of ranking (Method N) to each of 100 bootstrap samples is summarized in Table 8. Again a fair bit of variability of

Insert Tables 7 and 8 About Here

predictor ranks is exhibited in Table 8. It is clear, however, that P11 and P12 are two of the more important predictors, and that P5, P7, P8, and P10 are four of the least important predictors. Results of the 78 tests of the pairwise differences of predictor means for 100 bootstrap samples are given in Table 7. The three vertical lines next to the $Z_{(i)}$ values indicate the predictors for which the $Z_{(i)}$ means were not different. Again, it looks like P11 and P12 would, on the average, enhance predictive accuracy more than the other predictors, and that dropping P13 would actually increase the hit rate. Similar conclusions can be drawn from the application of ranking Method DZ and Method SZ. For this data set it is difficult to favor one of these two non-naive methods over the other.

Data Set TWO

Descriptive information for this data set is given in Table 4. The hit rates obtained by deleting each predictor in turn, $H_{(i)}$ values, and the transformed $Z_{(i)}$ values are given in Table 9. The predictor-rank counts based on the bootstrap samples are given in Table 10. This latter table reflects the considerable instability

Insert Tables 9 and 10 About Here

of the naive predictor ranks across repeated sampling. The results

of the 10 pairwise Student t tests are indicated in Table 9 with the vertical line next to the $Z_{(i)}$ values indicating nonsignificant differences among the $Z_{(i)}$ values for P1, P3, P2, and P4. This result is consistent with the predictor ordering according to Method SZ, a reasonable predictor ordering for this data set.

Comments

In any situation involving a design and some set of outcome measures on more than one variable, we have seen that a variety of different strategies may be used in an effort to determine those variables more "important" than others. Certainly others could be considered. Taken as a whole the examples which we have presented suggest but one compelling conclusion -- DON'T NAIVELY INTERPRET STRICT RANKINGS OF STATISTICS RELATED TO VARIABLE IMPORTANCE.

While we have based our own assessment of the various strategies on bootstrap results and, therefore, would seem thereby to be arguing for the superiority of such an approach, programs by which the bootstrap may be simply applied are not a feature of any statistical package. Furthermore, computational demands would seem to grow almost exponentially with the number of variables involved. The optimal solution probably lies in something like the more heuristic approaches considered here.

What we have not done is to address the issues of <u>when</u> it might make sense to undertake such an endeavor. What seems a reasonable desideratum on the issue of variable importance must rest with the nature of the particular outcome variables of

interest. If the outcome variables are conceptually indistinct from one another it would seem fruitless to undertake an effort to impose an ordering upon them. For example, if our outcome variables were represented by items composing a unidimensional scale, no real benefit could be obtained by exploring the issue of which is more important. In fact, in such a situation, the only statistic of any real interest would be the particular multivariate test itself (Wilks' lambda, Roy's largest root, etc.). Any results concerning the individual items themselves would be irrelevant.

On the other hand, if our outcome variables were to involve an index of socio-economic status, a measure of educational attainment, and a measure of the success in that educational attainment, then variable importance might indeed be a sensible question. One must of course still struggle with the complexities offered up by the interrelatedness of such constructs.

Some care should be taken in arriving at a variable ordering. One difficulty was addressed in this paper, namely, the assignment of ranks to the variables. Also, suppose it is concluded that one variable is relatively more important than the other variables being studied. It is only for the collection of variables studied where this can be said to be true. Changing the mix of variables involved could well change the conclusion.

A fairly recent effort on dealing with the problem of variable importance in descriptive discriminant analysis is that of Thomas (1989). The approach suggested by Thomas involves "combining" standardized discriminant function coefficients (SDFCs) and

structure coefficients (SCs) into something termed discriminant ratio coefficients (DRCs). A geometric rationale for the two DRCs is advanced. One DRC is simply the product of the SDFC and the SC for each variable. Using a DRC as a variable-ordering index is an approach yet to be appreciably studied.

In any event the central issue is not a statistical one. It rests squarely with the substantive questions for which answers are sought.

References

Blalock, H. M. (1961). Evaluating the relative importance of variables. American Sociological Review, 26, 866-874.

Eisenbeis, R. A., & Avery, R. B. (1972). Discriminant analysis and classification procedures. Lexington, MA: Heath.

Eisenbeis, R. A., Gilbert, G. G., & Avery, R. B. (1973). Investigating the relative importance of individual variables and variable subsets in discriminant analysis. Communications in Statistics, 2, 205-219.

Huberty, C. J (1972). Regression analysis and 2-group discriminant analysis. Journal of Experimental Education, 41, 39-41.

Huberty, C. J (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95, 156-171.

Huberty, C. J (1989). Problems with stepwise methods: Better alternatives. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J, Chou, T., & Benitez, E. B. (1990, April). The study of multivariate group contrasts. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Huberty, C. J, & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. Psychological Bulletin, 105, 302-308.

Huberty, C. J, & Smith, J. D. (1982). The study of effects in MANOVA. Multivariate Behavioral Research, 17, 417-432.

Huberty, C. J, & Wisenbaker, J. M. (in press). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2). Greenwich, CT: JAI Press.

Huberty, C. J, Wisenbaker, J. M., & Smith, J. C. (1987). Assessing predictive accuracy in discriminant analysis. Multivariate Behavioral Research, 22, 307-327.

James, M. (1985). Classification algorithms. New York: Wiley.

Karson, M. J., & Martell, T. F. (1980). On the interpretation of individual variables in multiple discriminant analysis. Journal of Financial and Quantitative Analysis, 15, 211-217.

Kruskal, W. (1984). Concepts of relative importance. Questiio, 8, 39-45.

Kruskal, W., & Majors, R. (1989). Concepts of relative importance in recent scientific literature. American Statistician, 43, 2-6.

McCabe, G. P. (1975). Computations for variable selection in discriminant analysis. Technometrics, 17, 103-109.

McKay, R. J., & Campbell, N. A. (1982). Variable selection techniques in discriminant analysis I. Description. British Journal of Mathematical and Statistical Psychology, 35, 1-29.

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. Journal of the American Statistical Association, 58, 275-309.

Mueller, R. O., & Cozad, J. B. (1988). Standardized discriminant coefficients: Which variance estimate is appropriate? Journal of Educational Statistics, 13, 313-318.

Pedhazur, E. J. (1982). Multiple regression in behavior research. New York: Holt, Rinehart and Winston.

Pruzek, R. M. (1990, April). Variable importance in common factor analysis. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Rencher, A. C. (1988). On the use of correlations to interpret canonical functions. Biometrika, 75, 363-365.

Schaafsma, W., & Van Vark, G. N. (1979). Classification and discrimination problems with applications, Part IIa. Statistica Neerlandica, 33, 91-126.

Tatsuoka, M. M. (1988). Multivariate analysis. New York: Macmillan.

Thomas, D. R. (1989). On the interpretation of discriminant function arising from a MANOVA test. Working Paper, WPS 89-16, School of Business, Carleton University, Ottawa.

Urbakh, V. Y. (1971). Linear discriminant analysis: Loss of discriminating power when a variable is omitted. Biometrics, 27, 531-534.

Van Vark, G. N. (1976). A critical evaluation of the application of multivariate statistical methods to the study of human populations from their skeletal remains. Homo, 27, 94-114.

Table 1

Group Sizes, Means, Standard Deviations[a], and Error Correlations for Data Set ONE

	Group 1	Group 2	Group 3	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
V1	2.90543	3.28741	3.33108	.4479	.5973	.5504	-.2900	.5596	.3717	.1486	.2648	.1811	-.2129	.0479	-.0488
	(.5043)	(.5052)	(.4906)												
V2	2.52000	2.85617	2.68676		.4259	.6741	-.2107	.5527	.2756	.5317	.2647	.3240	-.2062	-.0544	.1531
	(.6986)	(.7989)	(.7971)												
V3	2.89114	3.34321	3.20919			.5913	-.1988	.5062	.0517	.1802	.2229	.1188	-.2400	-.1054	-.0137
	(.4768)	(.5193)	(.5548)												
V4	2.65343	3.07395	2.95432				-.3360	.5107	.2457	.3695	.3196	.2906	-.3454	-.1514	-.0329
	(.5431)	(.5954)	(.6895)												
V5	3.17143	4.23457	4.97297					-.3544	-.2647	-.1820	.0216	-.0355	.2932	.2609	.0268
	(.9848)	(.7628)	(1.1177)												
V6	2.67629	3.20420	3.14757						.3898	.3189	.2398	.1961	-.0671	.0152	-.1650
	(.6901)	(.6389)	(.6654)												
V7	22.42857	23.80247	25.02703							.3600	.3319	.3981	.0732	.0531	.0356
	(2.0904)	(2.8610)	(2.5109)												
V8	23.65714	24.79012	24.27027								.3774	.5251	-.0343	-.0731	.2420
	(5.7340)	(4.9918)	(5.3471)												
V9	25.14286	26.59259	27.32432									.5878	.1279	.1414	.0591
	(3.9567)	(3.1416)	(3.6746)												
V10	23.31429	24.62963	26.05405										.2012	.0635	.1521
	(4.1570)	(4.5674)	(4.6064)												
V11	14.34286	19.27160	25.27027											.2112	-.1582
	(4.7832)	(3.6229)	(3.9627)												
V12	40.85714	52.87654	63.75676												-.3459
	(7.5504)	(7.2170)	(4.0098)												
V13	2.00000	1.39506	.48649												
	(2.1557)	(1.6022)	(.9894)												
n	35	81	37												

[a]Standard deviations in parentheses.

32

Table 2

F-to-Remove Values and Variable Rank-Orderings for Data Set ONE

	F-to-Remove		Rank Method N	Rank Method D	Rank Method S
V12	23.455		1	1	1
V11	16.483		2	2	2
V4	6.854		3	3	3
V5	5.067		4	4	4
V9	1.841	|	5	9	8
V3	1.731	|	6	9	8
V13	1.531	|	7	9	8
V7	1.358	|	8	9	8
V6	1.310	|	9	9	8
V2	1.166	|	10	9	8
V10	0.835	|	11	9	8
V8	0.197	|	12	9	12.5
V1	0.173	|	13	9	12.5

Note. See text regarding vertical lines next to the F values.

Table 3

Frequency of Method N F-to-Remove Ranks for Each Variable in 100 Bootstrap Samples from Data Set ONE

	Rank												
	1	2	3	4	5	6	7	8	9	10	11	12	13
V1	0	0	0	0	1	7	4	9	9	13	12	21	24
V2	0	0	0	4	9	8	19	15	15	3	15	7	5
V3	0	0	2	10	19	12	9	13	9	10	6	8	2
V4	0	8	49	26	6	7	0	2	0	2	0	0	0
V5	0	4	26	25	22	15	4	2	1	0	1	0	0
V6	0	0	2	5	7	9	13	8	12	12	9	13	10
V7	0	0	0	7	11	13	16	9	9	10	8	12	5
V8	0	0	1	0	1	3	5	7	9	16	17	19	22
V9	0	0	4	9	9	11	13	16	8	8	9	4	9
V10	0	0	4	7	3	4	8	7	16	17	10	12	12
V11	25	64	8	3	0	0	0	0	0	0	0	0	0
V12	75	24	1	0	0	0	0	0	0	0	0	0	0
V13	0	0	3	4	12	11	9	12	12	9	13	4	11

Table 4

Means, Standard Deviations, and Error Correlations for Data Set TWO

	Group 1	Group 2	Group 3	V2	V3	V4	V5
V1	16.80556	16.44444	14.00000	.0251	-.0567	-.0360	-.0393
	(5.2471)	(5.3084)	(4.5925)				
V2	16.77778	17.37037	15.21739		.5804	.6827	.7089
	(6.2295)	(5.6374)	(5.4352)				
V3	15.25000	15.25926	12.73913			.6253	.5062
	(6.2900)	(5.6745)	(4.5649)				
V4	17.13889	15.22222	14.21739				.7255
	(5.0774)	(5.8229)	(2.9993)				
V5	18.36111	17.81481	14,34783				
	(5.8415)	(3.3054)	(5.4239)				
n	36	27	23				

Table 5

F-to-Remove Values and Variable Rank-Orderings for Data Set TWO

	F-to-Remove	Rank Method N	Rank Method D	Rank Method 5
V5	3.685	1	2.5	1.5
V4	3.028	2	2.5	1.5
V1	2.500	3	2.5	3.5
V2	2.050	4	2.5	3.5
V3	0.743	5	5	5

Note. See text regarding vertical line next to the F values

Table 6

Frequency of Method N F-to-Remove Ranks for Each Variable in 100 Bootstrap Samples from Data Set TWO

	Rank				
	1	2	3	4	5
V1	20	23	20	23	14
V2	17	17	22	23	21
V3	44	4	21	31	40
V4	26	18	24	14	18
V5	33	38	13	9	7

Table 7

$H_{...}$ Values, $Z_{...}$ Values, and Predictor Rank-Orderings for Date Set ONE

	$H_{...}$	$Z_{...}$	Rank Method N	Rank Method DZ	Rank Method SZ
P12	.804	10.464	1	1.5	1
P11	.830	11.127	2	1.5	2
P4	.863	11.955	3.5	8	7.5
P1	.863	11.955	3.5	8	7.5
P2	.869	12.121	5.5	8	7.5
P6	.869	12.121	5.5	8	7.5
P7	.876	12.287	8.5	8	7.5
P8	.876	12.287	8.5	8	7.5
P9	.876	12.287	8.5	8	7.5
P10	.876	12.287	8.5	8	7.5
P5	.882	12.452	11.5	8	7.5
P3	.882	12.452	11.5	8	7.5
P13	.889	12.618	13	8	13

Note. See text regarding vertical lines next to the $Z_{...}$ values.

Table 8

Frequency of Method N Hit Rate Ranks for Each Prediction in 100 Bootstrap Samples from Data Set ONE

	Rank												
	1	2	3	4	5	6	7	8	9	10	11	12	13
P1	0	3	4	8	9	7	16	11	16	7	13	6	0
P2	0	1	9	8	11	7	7	8	19	7	8	7	9
P3	1	7	12	9	6	10	9	12	9	6	8	9	2
P4	5	11	26	11	8	10	4	5	4	4	5	5	2
P5	0	5	8	10	6	12	11	16	7	5	8	16	6
P6	0	5	6	9	12	8	12	10	10	9	10	5	4
P7	1	1	4	7	9	4	7	9	9	12	13	11	13
P8	0	0	2	3	10	4	13	14	13	6	15	14	6
P9	1	4	11	12	10	8	10	5	9	10	7	7	6
P10	0	1	2	4	7	6	9	13	7	15	11	20	5
P11	38	37	13	3	3	1	2	1	0	0	1	1	0
P12	58	24	9	3	3	0	2	1	0	0	0	0	0
P13	0	4	5	8	11	11	13	14	9	10	4	9	2

Table 9

$H_{(i)}$ Values, $Z_{(i)}$ Values, and Predictor Rank-Orderings for Data Set TWO

	$H_{(i)}$	$Z_{(i)}$	Rank Method N	Rank Method DZ	Rank Method SZ
P5	.442	1.883	1	3	1
P1	.477	2.563	2.5	3	3.5
P3	.477	2.563	2.5	3	3.5
P2	.488	2.790	4.5	3	3.5
P4	.488	2.790	4.5	3	3.5

Note. See text regarding vertical lines next to the $Z_{(i)}$ values.

Table 10

Frequency of Method N Hit Rates for Each Predictor in 100 Bootstrap Samples from Data Set TWO

	Rank*				
	1	2	3	4	5
P1	14	22	24	22	18
P2	21	18	14	18	29
P3	16	27	21	24	12
P4	22	24	18	17	19
P5	8	14	23	26	29

* A number of tied hit rates resulted. A count for a rank of "1" includes those samples where a predictor, when deleted, yielded the smallest hit rate or tied for the smallest hit rate. A count for a rank of "2" includes those samples where a predictor was assigned a rank of 2 or of 2.5. And so forth for the remaining three ranks.